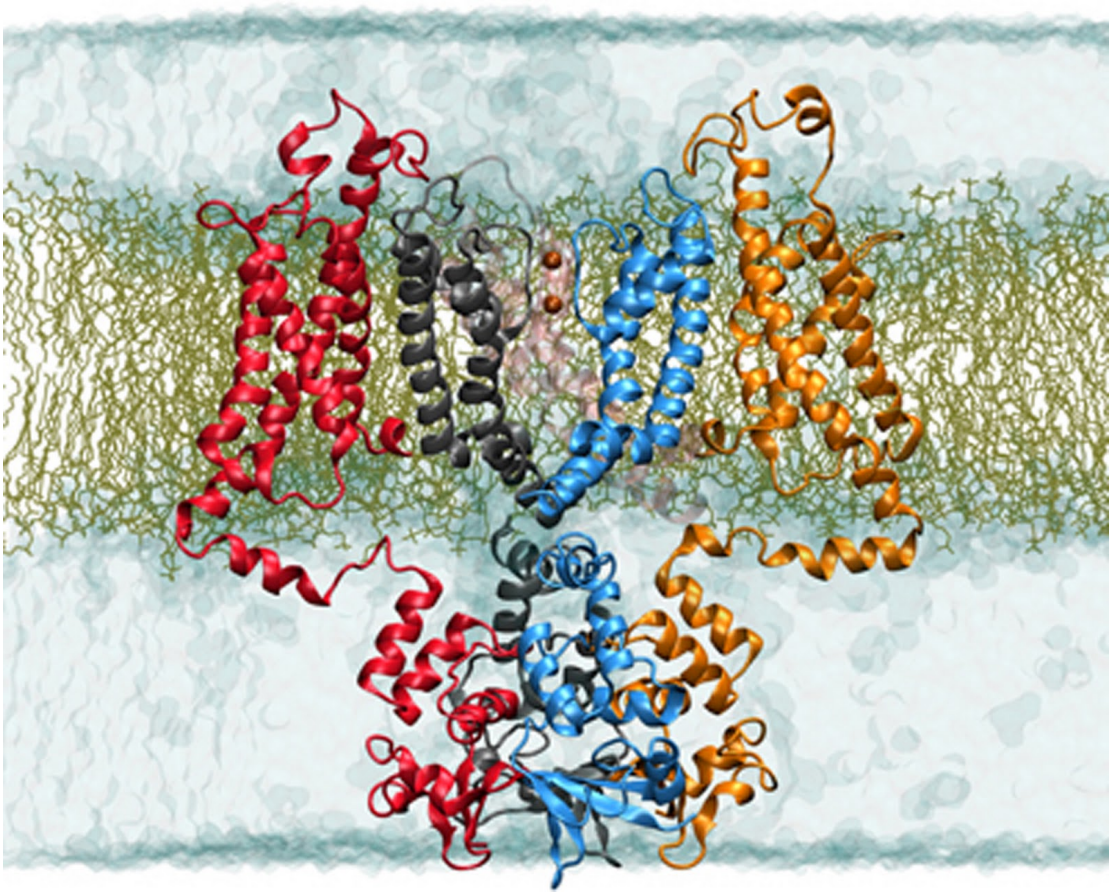
The use of solution NMR labeling techniques in solving structures of membrane proteins by solid state in-cell NMR.



By: Niels Remers
3219860
Supervisor: Dr. Gert Folkers
Second supervisor: Prof. Dr. J. Antoinette Killian

Table of content:

List of figures:

1. Introduction

Solving high-resolution membrane protein structures has proven to be a challenge. Among the large number of known protein structures, membrane protein structures are underrepresented. Despite the fact that membrane proteins are very promising drug targets[1] and by estimate 15% to 39% of the human genome is dedicated to membrane proteins[2,3], only a small number of membrane protein structures is solved[4].

X-ray crystallography and solution state nuclear magnetic resonance (NMR), the two major conventional approaches of solving protein structures, are not well suited to study membrane protein structures. Important for the integrity of a membrane protein are the surrounding lipids[5,6]. These lipids complicate the growth of high quality crystals and rapidly reorienting NMR samples. Furthermore the expression and subsequent purification of membrane proteins is far from straightforward.

Expression and purification of membrane proteins requires additional complicating steps. Whereas most cytoplasmic proteins already are soluble, thus ready to be purified, membrane proteins need to be extracted from the membrane before they can be purified. The membrane protein needs to be solubilized and maintain solubility during the purification process before it can be reconstituted in an artificial lipid bilayer[7,8]. The solubilization of membrane proteins is difficult because of the hydrophobic trans-membrane domains of membrane proteins. The solubilization and reconstitution are rate- and yield-limiting steps in the purification of membrane proteins. These steps can be avoided by using in-cell measurements.

In contrast to solution state NMR, solid state NMR does not require fast tumbling proteins for measurements. This makes solid state NMR a technique well suited to study membrane proteins, fibril proteins, and large protein complexes[9]. Membrane bound proteins which do not have a tumbling rate necessary for solution state NMR[10,11] can be studied as well by solid state (in-cell) NMR.

Solid state in-cell NMR is an excellent technique to study membrane proteins[12,13]. The in-cell aspect resolves the requirement of the solubilization and reconstitution steps, resulting in a far more straightforward expression scheme. And the rate of tumbling is not important for solid state NMR.

In this study different labeling schemes are investigated for their use for structural assignment of membrane proteins by in-cell solid state NMR. By combining clever labeling schemes with in-cell solid state NMR, this technique can prove to be very useful in filling in the large gap of structural data for membrane proteins.

2. NMR labelling techniques

Studying proteins by NMR requires the proteins to be labelled. Stable isotopes such as ^{13}C, ^{15}N, and ^{2}H are detectable by NMR and need to be incorporated into the protein of interest. However large proteins may suffer from spectral crowding which renders unresolved peaks unidentifiable. Clever use of labelling techniques reduces the spectral crowding and improves the peak assignment and subsequent structure determination.

This study focuses on solution state NMR labelling techniques for their use in in-cell solid state NMR. A great number of different labelling techniques is available such as reduced labelling[14], methyl group labelling[15], reversed labelling[16], and segmental isotope labelling[17]. These will be discussed in this chapter.

Doing in-cell NMR requires the labelling technique to selectively label the protein of interest. This is in contrast to classical NMR where the protein is purified from the cells and background labelling is not relevant. Selective labelling of the protein of interest will be the first issue discussed in this chapter.

2.1 Selective labelling of a protein of interest

Methods for expressing labelled proteins are well established. However after expression most proteins are purified, rendering selectivity of a labelling technique irrelevant. For in-cell NMR selectivity of a labelling technique is more important because the cells for the protein expression are subsequently used in the data acquisition. Any a-specifically labelled protein leads to background signals, potentially over-crowding the NMR spectra with background signal.

Special expression schemes were designed to minimize the background signal. A normal expression scheme is divided in two parts; first cells are grown to an optimal density, and second the expression of the protein of interest is induced at the optimal density. Normally the expression of bacterial proteins continues during the expression of the protein of interest. By blocking the expression of bacterial proteins during the expression of the protein of interest one can specifically label the protein of interest.

Rifampicin is a drug that inhibits the bacterial RNA polymerase and stops the bacterial protein expression. Rifampicin does not block the T7 bacteriophage RNA polymerase[18] which is regularly used for protein over-expression. Using rifampicin stops the bacterial protein expression. After the bacterial protein expression is stopped, IPTG can be used to start the expression of the protein of interest, using rifampicin in conjunction with IPTG effectively stops the bacterial protein expression during the expression of the protein of interest.

Before rifampicin and IPTG are added the bacterial cells are grown similar to a normal expression scheme in unlabelled medium. This prevents a-specific labelling of endogenous proteins. At the optimal cell density the cells are harvested by centrifugation and re-suspended in medium containing IPTG,

rifampicin, and the desired label. The expression of the protein of interest is induced by IPTG, the expression of endogenous bacterial proteins is blocked by rifampicin, and the protein of interest is labelled with the label of choice.

In 2001 the first spectrum of an in-cell labelled protein was recorded inside the living cells[19]. The expression scheme was further investigated. The effects of adding rifampicin, refreshing the medium, and overexpression were mapped[20]. This study showed that the effect of Rifampicin on the background levels is negligible when the protein of interest has an over-expression level of ~5% or higher. Therefore rifampicin can be used for difficult, low-expressing proteins which are not sufficiently over-expressed[21].

Rifampicin is also used to prevent cell-death during the data acquisition process[22]. This study shows that there is more than one cause for cell-death during data acquisition. Rifampicin is effective against the late stage cell-death but promotes the early stage cell-death caused by the MazEF module. The MazEF module is a toxin-antitoxin system, which is constantly expressed. MazE counters the toxic effects of MazF. Under stress, or by adding rifampicin, the expression of the MazEF module can be inhibited. Because the antitoxin MazE is short lived, the lytic effects of MazF prevail. Therefore this study suggests the use of an *E. Coli* strain that lacks the MazEF module in conjunction with rifampicin for future in-cell NMR measurements.

2.2 Reduced labelling

2.2.1 Reduced labelling by labelled carbon sources

Solid-state NMR measurements can be made more efficient by selective and extensive ^{13}C labelling. Selective and extensive ^{13}C labelling reduces the spectral crowding and facilitates the sequence specific assignment.
One way to reduced label proteins is by making use of the metabolism and amino acid synthesis pathways of the *E. Coli*. The metabolism and amino acid synthesis pathways convert glucose, glycerol or any other available carbon source into amino acids; hence a reduced labelled carbon source results in reduced labelled amino acids. The existing knowledge of the metabolism and amino acid synthesis pathways can be used to devise intelligent labelling schemes. The amino acids are synthesized from various intermediates of the *E. Coli* metabolism (fig 1).
The metabolism can be divided in two sections, the glycolysis and the citric acid cycle (CAC). The glycolysis is the linear part and amino acids derived from these intermediates are synthesized using the pentose phosphate pathway (PPP). Because of the linear character of the glycolysis the labelling of these amino acids occurs highly efficient and the position of the label is conserved[14].
In contrast the amino acids derived from the CAC are labelled in a non-random mixture (fig S1). This is because of the cyclic nature of the CAC which spreads a single label to multiple positions.
For example when 2-glycerol is used as an carbon source 3/5 of the CO positions of Proline are labelled and only 1/5 of the Cα positions of Proline are labelled. Although Proline is not uniformly labelled, the label still has a non-random distribution that can be used in the structure determination.

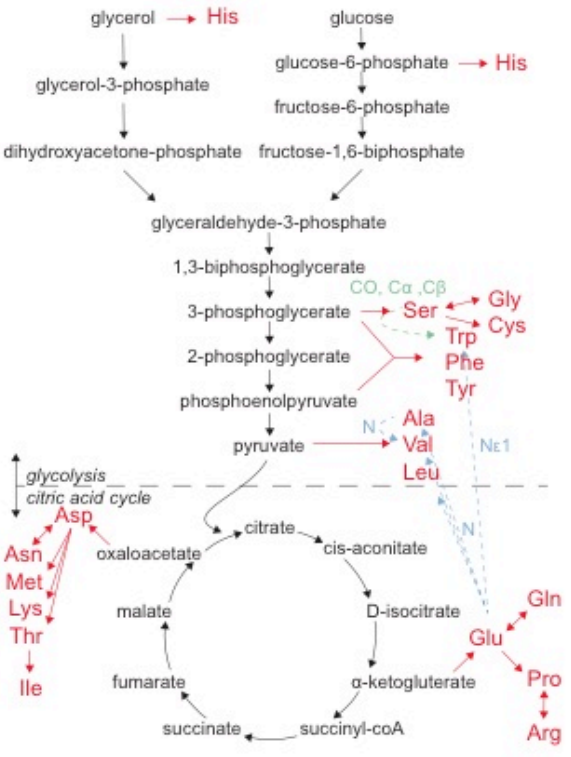


Fig 1. *E. Coli* metabolism. Consisting of the glycolysis and the citric acid cycle. Amino acids synthesized are depicted in red. Nitrogen scrambling is depicted by blue arrows. Figure adapted from M. Hong et al. (1999)[23]

The main carbon sources used for reduced labelling are glucose, glycerol, succinic acid, and acetate. Glycerol and glucose are the natural start of the glycolysis pathway. Most amino acids are derived from metabolic intermediates downstream of the glyceraldehyde-3-phosphate, hence the similarities between the labelling schemes for [1,3-^{13}C]-glycerol, [1-^{13}C]-glucose, and [2-^{13}C]-glycerol, [2-^{13}C]-glucose (table 1 and fig. 2)[24].
However glucose labelling is only half as efficient as glycerol labelling since glucose breaks up into two glyceraldehyde molecules of which only one contains the label (fig 2).

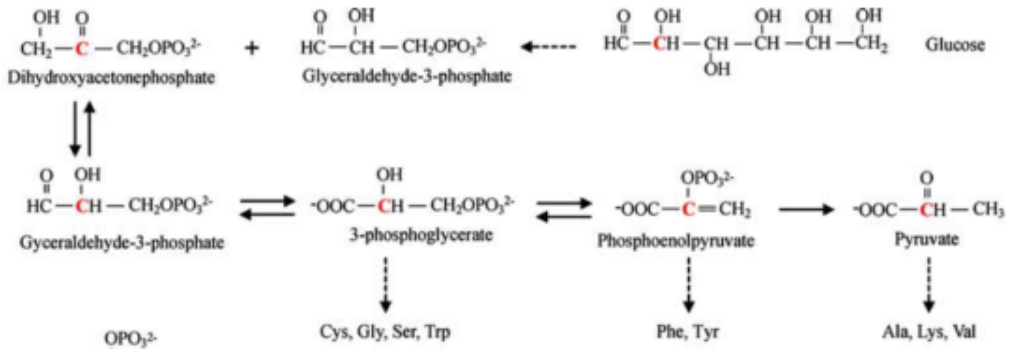


Fig 2. Glucose labelling is half as efficient as glycerol labelling. In this figure the second carbon atom is labelled red and its fate can be followed. During glycolysis glucose is converted into glyceraldehyde-3-phosphate and dihydroxyacetone phosphate. Dihydroxyacetone phosphate is quickly converted into glyceraldehyde-3-phosphate. Only one of the two resulting glyceraldehyde-3-phosphates is labelled. Figure from Lundström et al. (2007)[24]

Next to using glycerol or glucose as the sole carbon source other carbon sources can be used, succinate (citric acid cycle) can be used to label all amino acids except Ile and Leu. Ile and Leu should be added to the medium if succinate is used as the carbon source. [1,4-^{13}C]-succinate or [2,3-^{13}C]-succinate labelling results in complementary labelling schemes (table 1 and fig 3)[14,25].
Ile and Leu can be added labelled or non-labelled to the medium, reducing the 2D correlations significantly[26]. Acetate labelling results in a similar labelling scheme as succinate labelling, though with a lower labelling frequency[27].

To evade the non-random labelling mixtures for the amino acids derived from the CAC a new labelling protocol based on the glucose and glycerol labelling protocol was designed. This labelling protocol is dubbed TEASE which stands for *t*en-*a*mino acid *s*elective and *e*xtensive labelling. Next to the labelled glycerol (or glucose), unlabelled amino acid products from the citric acid cycle are added in the medium. Hence only the amino acid products from the glycolysis and pentose phosphate pathway (PPP) are labelled and the scrambling is reduced to a minimum. This labelling protocol is extremely relevant for the study of membrane proteins because the membrane embedded sections of these proteins are enriched in labelling since the hydrophobic amino acids are mostly synthesized via the glycolysis and PPP[28].

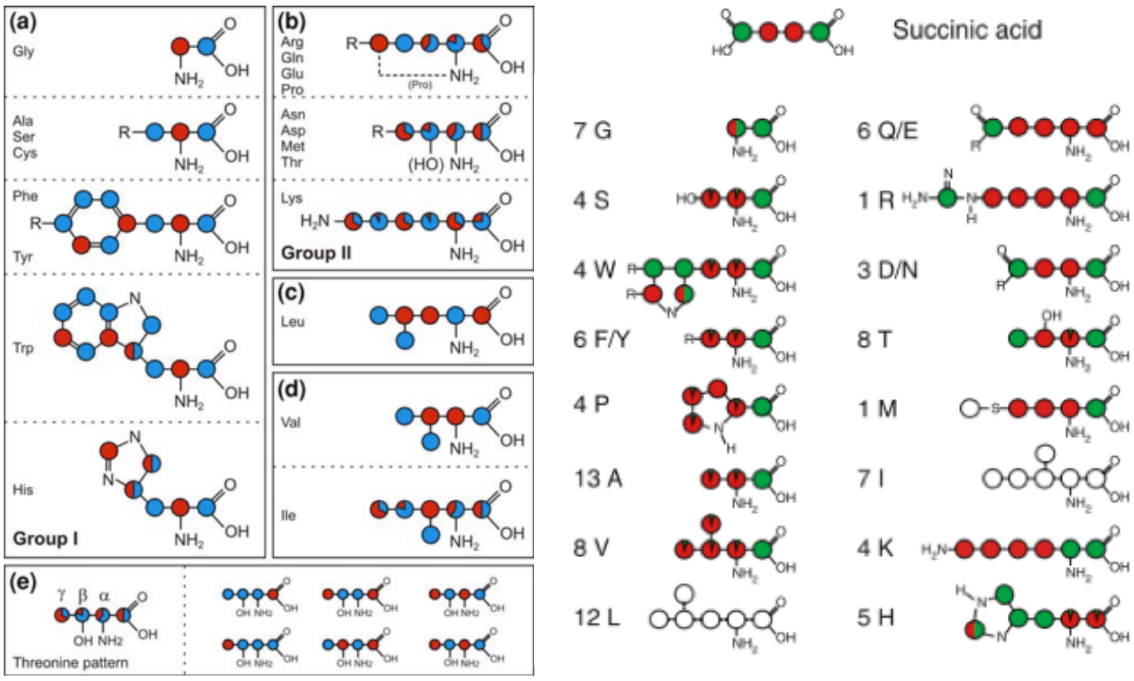


Fig 3. ^{13}C Labelling patterns for glycerol and succinate. Right [1,3-^{13C}]-glycerol in blue and [2-^{13C}]-glycerol in red. Left [1,4-^{13C}]-succinate in green and [2,3-^{13C}]-succinate in red. Glycerol labelling figure from Higman et al. (2009)[14] and succinate labelling figure from van Gammeren et al. (2004)[25]

Table 1. Labelling schemes of 1-glucose, 2-glucose, 1,3-glycerol, 2 glycerol, 1,4-succinate, and 2,3-succinate. Amino acids are either ~100% labelled, ~50% labelled or ~25% labelled. [1]residue labelling only predicted.

	1-glucose	2-glucose	1,3-glycerol	2-glycerol	1,4-succinate	2,3-succinate
	E. Coli	*E. Coli*	*E. Coli*	*E. Coli*	*R. Acidophila*	*R. Acidophila*
Gly	-	α	CO	α	CO, α	α
Cys	β[1]	α	CO, β	α[1]	-	-
Ser	β	α	CO, β	α	CO	α, β
Ala	β	α	CO, β	α	CO	α, β
Val	γ1, γ2	α, β	CO, γ1, γ2	α, β	CO	α, β, γ1, γ2
Leu	α, δ1, δ2	CO, β, γ	α, δ1, δ2	CO[1], β, γ	-	-
His	CO, δ2	α, γ	CO, β, γ, δ	α[1], γ, δ2, ε1	γ, ε1	CO, α, δ2, ε1
Trp	β[1], δ1[1], δ2[1]	α, γ, δ1, δ2, ξ1, ξ2	CO, β, γ, δ1, ε2, ξ2, η2	α, γ, δ2, ξ3	CO, γ, δ1, δ2	α, β, δ1, ε2
Phe	β, γ, δ1, δ2	α, γ, ε1, ε2	CO, β, δ1, δ2, ε1, ξ	α, γ, ε2	CO	α, β
Tyr	β, δ1, δ2, ξ	α, γ, ε1, ε2	CO, β, δ1, δ2, ε1, ξ	α, γ, ε1, ξ	CO	α, β
Glu	CO[1], α, β, γ	α, δ	CO, α, β, γ	CO[1], α, β	δ	CO, α, β, γ
Gln	CO[1], α, β, γ	α, δ	CO, α, β, γ	CO[1], α, β	δ	CO, α, β, γ
Pro	CO[1], α, β, γ	α, δ	CO, α, β, γ	CO[1], α, β, δ	CO	α, β, γ, δ
Arg	CO[1], α, γ	α, δ, ε	CO, α, β, γ	CO[1], α, β, δ	CO, ε	α, β, γ, δ
Asp	CO[1], α, β, γ[1]	α, γ	CO, α, β, γ	CO[1], α, β, γ	CO, γ	α, β
Asn	CO[1], α, β, γ[1]	α, γ	CO, α, β, γ	CO[1], α, β, γ	CO, γ	α, β
Met	CO[1], α, β, γ[1]	α, γ	CO, α, β, γ	CO[1], α, β[1], γ, ε	CO, γ	α, β
Thr	CO[1], α, β, γ[1]	α, γ	CO, α, β, γ	CO[1], α, β, γ	CO, γ	α, β
Lys	β, γ[1], δ, ε	α, γ, ε	CO, α, β, γ, δ, ε	CO, α, γ, δ, ε	CO	-
Ile	CO[1], α, γ1, γ2, δ1	α, β, δ	CO, α, γ1, γ2, δ	CO[1], α, β, γ1, δ1	-	-
reference:	Hong[23]	Lundström et al.[24]	Higman et al.[14]	Higman et al.[14] & Hong[23]	Gammeren et al.[25]	Gammeren et al.[25]

Table composed with results from several studies[14,23–25].

2.2.2 Fractional labelling

Fractional labelling is a method in which a label is added only fractionally. Fractional ^{13}C labelling between 10% and 100% ^{13}C label has been done[29]. The fractional ^{13}C labelling is achieved by supplementing the growth medium with both $^{12}C_6$-glucose and $^{13}C_6$-glucose. By adding less ^{13}C label the chances of non-relevant cross-peaks reduce significantly, however the signal to noise ratio becomes less favourable.

Fractional ^{13}C labelling has also been used for stereospecific assignments of Val and Leu residues[30]. In this study 10% ^{13}C and 90% ^{12}C was used. Because of the fractional ^{13}C labelling there are two types of adjacent ^{13}C labelled pairs; random- and non-random-pairs. Because the ^{13}C incorporation in this study was 14% chances of a random adjacent labelled ^{13}C pair were 2% (14% times 14%). A non-random pair occurred with a yield of 14%. Therefore only one strong cross peak connects the Cβ with a Cγ and there is no cross peak connecting the second Cγ. The cross peak originates from the non-random ^{13}C pair and can be used for stereospecific assignments.

Fractional deuteration is a powerful technique that improves several spectral issues. By fractionally deuteration the hydrogen correlation experiments improve on spectral dispersion and reduce in line width. Furthermore, because the residual hydrogens are preferably found on the side chains, the structural determination of these is facilitated[31].

2.2.3 Methyl group labelling

Methyl groups are useful reporters of protein structure. Several chemical compounds can be used to specifically label methyl groups. Pyruvate, 2-ketoisovalerate, and 2-ocobutanoate are used for this purpose. These pyruvate-like ketogluterates are used to specifically label methyl groups of Ile, Leu, Val residues (table 2).

Table 2. Methyl labelling patterns

Methionine labeling				
	E. Coli	*E. Coli*	*E. Coli*	*E. Coli*
	3-pyruvate	2-ketoisovalerate	2-oxobutanoate	2-oxaloacetate
Ala	β			
Val	γ1, γ2	Whole		
Leu	δ1, δ2	Sidechain		
Glu				
Gln				
Pro				
Arg				
Asp				Whole
Asn				Whole
Met	ε			Co, α, β, γ
Thr				Whole
Lys				Co, α, β, γ
Ile	γ2		CO, α, γ2, δ	
reference:	Mulder et al.[33]	Sibille et al.[32]	Sibille et al.[32]	

Table composed with the results from several studies[32,33.]

The conversion of 2-ketoisovalerate into Val requires a single step. Hence the labelling scheme of the precursor is entirely preserved in the Val residue. The conversion of 2-ketoisovalerate into Leu requires acetyl CoA, which is incorporated in the CO and Cα positions. The labelling of the CO and Cα positions depends on the added carbon source in the medium, the sidechain C-positions have the same labelling scheme as the precursor.

The biosynthesis of Ile from 2-oxobutanoate is even more complex, both the Cβ and Cγ1 come from the carbon source in the medium but the rest of the C positions have the same labelling pattern as the precursor[32].

2-keto-isovalerate and 2-oxobutanoate result in unscrambled labeling of Ile, Leu, and Val residues, in contrast: adding labeled Ile, Leu and Val results in scrambling. These precursors can be used to selectively $^{13}CH_3$ label a single

methyl group of Ile, Leu, and Val. The backbone carbons of these residues can be labelled or non labelled by adding labelled or non labelled glucose to the medium[15,32].

3[$^{13}CD_2H$]-pyruvate in D_2O as the sole carbon source results in labelled methyl groups of Ala (β), Val (γ), Leu (δ), and Ile (γ2)[33]. This labelling technique enables ^{1}H detected experiments for ssNMR in a resolution comparable with solution state ^{1}H, ^{13}C experiments[34].

The methyl groups of Ala[35,36], Met[37], and Thr[38] can be specifically labelled by adding these amino acids to the medium. Scrambling of Ala to Val or Thr to Ile can be prevented, if undesired, by adding Val or Ile to the medium[36].

2.2.4 Reduced labelling by labelled amino acids

Labelling a single amino acid is theoretically a great way to reduce signal overlap. The largest drawback of this method is scrambling. However scrambling is a non-random process, which can be used in the structure calculations or countered to prevent any scrambling.

Adding a single labelled amino acid type to the growth medium often results in the incorporation of a multitude of labelled amino acids, this is due to the metabolic processes occurring in the *E. Coli*. To understand which atom of which amino acid scrambles to which position is important for labelling purposes. This knowledge can be used in the structure calculations or to prevent scrambling. The amino acids can be subdivided categorized to the chemical compounds they are derived from. There are 5 (6) subclasses: the 3-phosphoglycerates (Ser, Cis, and Gly), the phosphoenolpyruvates (Phe, Tyr, and Trp), the pyruvates (Ala, Val, and Leu), the α-ketogluterates (Glu, Gln, Arg, and Pro), the oxaloacetates (Glu, Gln, Arg, and Pro), and finally the ribose-5-phosphate (His). It is important to realize that some of the amino acids within these families are precursors for the other amino acids within the family. For example, Asp is a precursor for the entire oxaloacetates family. Therefor labelling these amino acids can result in scrambling of the label. Furthermore Glu is a common precursor for amide groups of the pyruvates.

Scrambling for each position of each amino acid is different, while most ^{13}C-CO positions scramble minimally, backbone Nitrogen positions scramble with higher rates[39]. Amino acids such as Gly, Glu, and Gln ^{13}C labels scramble in a non-random fashion, which can be used in the structure calculations. Gly is conversed into Ser which is subsequently conversed into Trp. Glu and Gln are interconvertible and Glu is the precursor for Arg and Pro (fig 4)[39]. (fig S2 and S3)

When undesired the scrambling can be supressed in multiple ways. The use of auxotrophic cell-lines or manipulating the product feedback inhibitory loops of the metabolic pathways are options to supress scrambling[40–42]. Auxotrophic cell lines are cell lines unable to synthesize essential amino acids. The enzymatic reactions, which cause the scrambling, are inactivated in these cell lines. Most auxotrophic cell lines are auxotroph for one amino acid, however cell lines auxotrophic for pairs of amino acids are also available[40].

Because the use of auxotrophic cell lines can lead to sub optimal protein production due to low grow or expression rates other methods of preventing scrambling have been investigated[42]. A combination of enzyme inhibitors (especially the Glu/Gln conferting enzymes) and additional non-labelled amino acids supplemented in the growth medium supressed scrambling effectively.

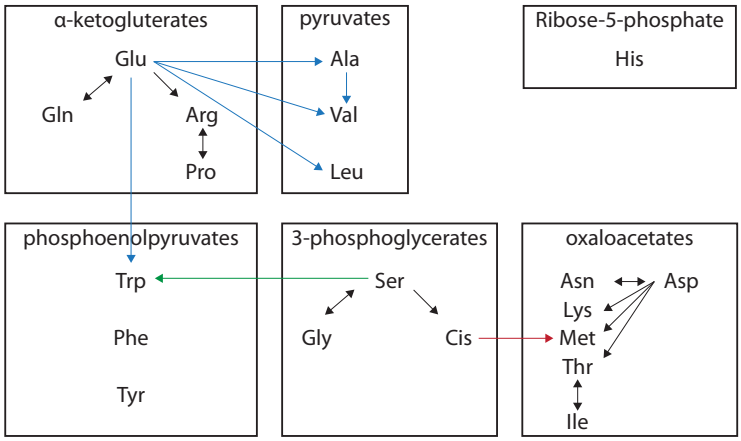


Fig 4. Amino acid scrambling. The amino acids can be subdivided into six families based on their common precursor. Black arrows indicate scrambling of the entire amino acids. Blue arrows indicate a transfer of the backbone amino group (from Glu to Trp Nε is transferred). The green arrow indicates the transfer of the CO, Cα, and Cβ. The red arrow indicates the transfer of the sulphur atom. Figure composed on basis of several articles.

2.2.5 Reversed labelling

Reversed labelling is a technique very similar to adding labelled amino acids. The major difference is the adding of unlabelled amino acids in a background of uniform labelling. “unlabels” used for reversed labelling include ^{12}C and ^{14}N. The aim is to reduce spectral crowding. In principle reversed labelling has similar limitations as single amino acid labelling; scrambling of amino acids has to be taken into account or prevented. A major difference between label scrambling and unlabel scrambling is the relative low scrambling of ^{14}N unlabel of Gln[43]. Because Leu, Phe, Tyr, Val, Lys, and Trp do not scramble significantly they have been used in reversed labelling studies[16,44,45].

Reversed labelling and selective amino acid labelling can be used in a combinational fashion[46]. The combination of labels in a set of samples results in a different labelling pattern for each amino acid across these samples. The aim of combinational techniques is to use as few samples as possible whilst maintaining unique labelling patterns. When using only one label, such as ^{15}N, only four samples are required to result in a unique labelling pattern for the twenty amino acids. When using multiple labels even less samples are needed to uniquely cover the twenty amino acids[47].

2.2.6 Single position labelling

Single position labelling can be used to greatly reduce spectral crowding and even to incorporate unnatural amino acids. In *E. Coli* this is mainly done by using the Amber stop codon (TAG) since 93% of stop codons in *E. Coli* are either ochre- or opal-stop codons (TAA and TGA). Using the Amber stop codon results in the least impact for the *E. Coli* proteome[48]. In principle any residue can be replaced by an unnatural amino acid, although the structural consequences of this replacement should be taken into account[49]. The variety of unnatural amino acids that can be incorporated is enormous[50]. A great number of structurally different unnatural amino acids already have been used in previous studies[51].

2.2.7 Segmental isotope labelling

Segmental isotope labelling is a technique for reduced labelling of proteins. Segmental isotope labelling only labels a single domain or part of a protein. Segmental isotope labelling sample preparation can be divided in three categories: Native Chemical Ligation (NTL), Expressed Protein Ligation (EPL), and Protein Trans Splicing (PTS). NTL is not of relevance for this study since it involves the preparation of protein using solid phase peptide synthesis. EPL, which requires specific in-vitro ligation conditions, is also not of relevance for this study[52]. PTS is the only segmental isotope labelling strategy that can be used to express and ligate the protein fragments in-vivo[53].

Protein trans splicing is a technique in which two or more protein domains are expressed separately. This separation could be achieved by the use of two different promoters. Both protein domains are expressed with a terminal intein fragment. Upon reconstitution of the two intein fragments, the intein is autocatalytically cleaved and the two protein domains reunited. Disadvantage of the technique is the requirement of a cysteine at the protein splicing site and general low yields because of the overall complexity of the trans splicing system. However recently a three fragment ligation has been preformed[17], showing the great possibilities of this technique.

2.2.8 Silac.

Stable isotope labelling is also used in mass spectrometry. The main use in mass spectrometry is to introduce a specific heavy amino acid in one cell culture and to compare resulting heavy peptides with normal peptides. This method is called SILAC, which stands for stable isotope labelling by amino acids in cell culture[54]. Most commonly used “heavy” amino acids are Leu-δD3[55], Lys-γδD4[56], Met-ε^{13}CD3[57], Arg-^{13}C6[58], and Tyr-^{13}C9[59]. No significant scrambling of the labels is reported except for Arg which is converted into Pro when there is an excess of Arg, vice versa Pro is converted into Arg when there is an Arg shortage. The ideal amount of added labelled Arg to the medium can be easily determined[54]. SILAC is used in HeLa or other mammalian cells. However scrambling rates can be expected to be similar in *E. Coli* since Leu, Lys, and Met are no precursors for other amino acids. Furthermore Arg and Pro are also interconvertible in *E. Coli*.

3. Protein detection schemes

For solid state MAS NMR different protein detection schemes are described in literature. A short list of these possible protein detection schemes follows.

3.1 PDSD and DARR

PDSD stands for proton driven spin diffusion, it is usually the first spectrum to record for a solid state experiment. The magnetization is transferred from hydrogen atoms to ^{13}C nuclei. From there the magnetization is transferred to all nearby ^{13}C nuclei (fig 5). At shorter mixing times all carbon nuclei within a residue are coupled, at longer mixing times also couplings between residues will appear[60]. The DARR scheme is very similar to the PDSD scheme, however it is more efficient at longer mixing times[61].

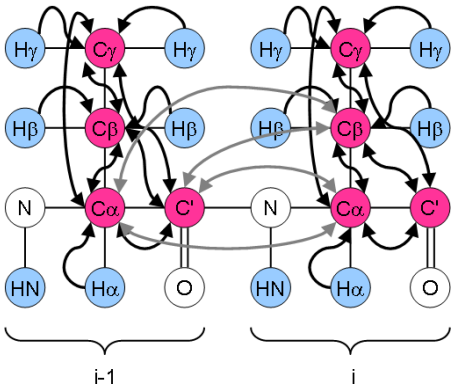


Fig 5. Observed magnetization transfers for a PDSD and DARR protein detection scheme. Nuclei observed are in pink, nuclei through which magnetization flows are in light blue. Magnetization transfer at short mixing times in black arrows and at longer mixing times in grey arrows. Figure from www.protein-nmr.org.uk (10-2012).

3.2 NCA and NCO

Both NCA and NCO are two-dimensional protein detection schemes, therefore they tend to be crowded for larger proteins. In both the schemes the magnetization is transferred from the NH, via the N (fig 6). Whereas the NCA experiment the magnetization remains within the residue, with the NCO experiment sequential links can be observed[62].

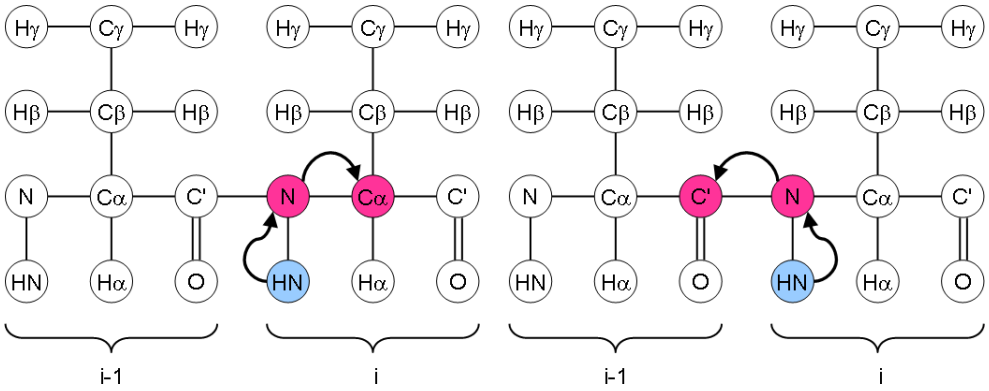


Fig 6. Observed magnetization transfers for a NCA and NCO protein detection scheme. Nuclei observed are in pink, nuclei through which magnetization flows are in light blue. Magnetization transfer is in black arrows. Figure from www.protein-nmr.org.uk (10-2012).

3.3 NCACX and NCOCX

The NCACX and NCOCX are basically the NCA and NCO protein detection schemes with an added PDSD or DARR step[63]. This makes these schemes three-dimensional, the chemical shift is evolved on the N, CA or CO, and detected on other ^{13}C nuclei (fig. 7). Because these schemes are three dimensional they suffer less from spectral crowding. A two-dimensional version is also possible where the evolution time on the CA or CO nuclei is left out. As with the original PDSD and DARR schemes the magnetization time can be shorter (interresidual) and longer (intraresidual).

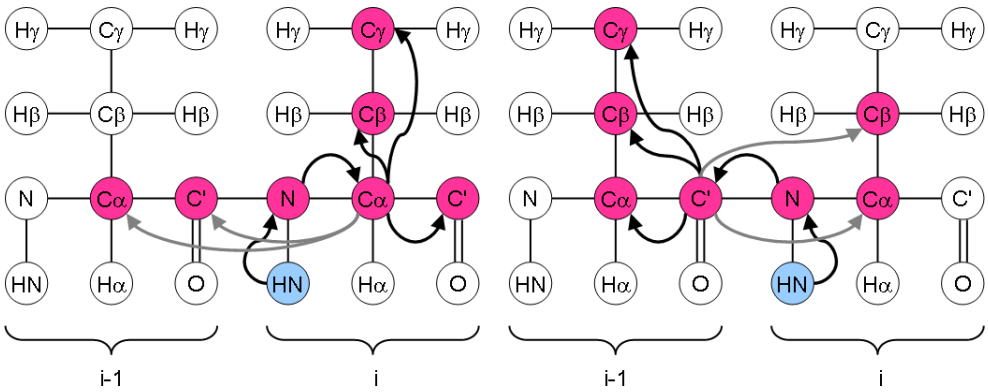


Fig 7. Observed magnetization transfers for a NCACX and NCOCX protein detection scheme. Nuclei observed are in pink, nuclei through which magnetization flows are in light blue. Magnetization transfer at short mixing times in black arrows and at longer mixing times in grey arrows. Figure from www.protein-nmr.org.uk (10-2012).

3.4 NCACB

The NCACB protein detection scheme is a variant of the NCACX scheme. Instead of a PDSD or a DARR, a DREAM (dipolar recoupling enhanced by amplitude modulation)[64] step is added after the NCA scheme, thereby specifically transferring the magnetization to the CB (fig 8). The NCACB protein detection scheme can be used for decrowding a NCACX system. Although the DREAM step is optimised for transfer to CB atoms, CY atoms will also become exited to some extend. However because the DREAM step is a double quantum step the CB peaks will be negative and the CY peaks positive, allowing easy identification[63].

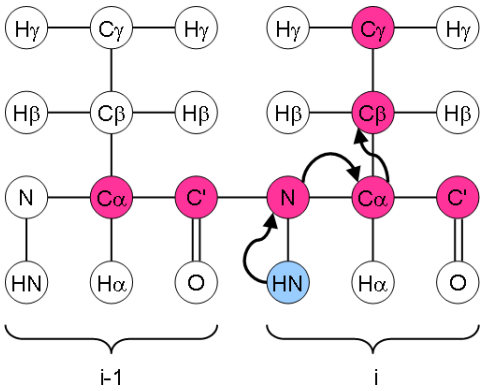


Fig 8. Observed magnetization transfers for a NCACB protein detection scheme. Nuclei observed are in pink, nuclei through which magnetization flows are in light blue. Magnetization transfer is in black arrows. Figure from www.protein-nmr.org.uk (10-2012).

3.5 CANCO and CANCOCX

The CANCO[65] and CANCOCX[66] protein detection schemes can provide additional sequential assignments to the NCACX and NCOCX schemes. The magnetization is transferred in three cross polarization steps from the hydrogen, via the CA and the N, to the CO (fig 9). Due to the three cross polarization steps this scheme suffers from a low signal to noise ratio. Making use of a PDSD step after the CANCO scheme makes the CANCOCX a four-dimensional scheme. Because of the four-dimensional character of the protein detection scheme, signal overlap is decreased. Both the CANCO and the CANCOCX can be lowered in dimensionality by omitting the N (and/or CO for CANCOCX) evolution time.

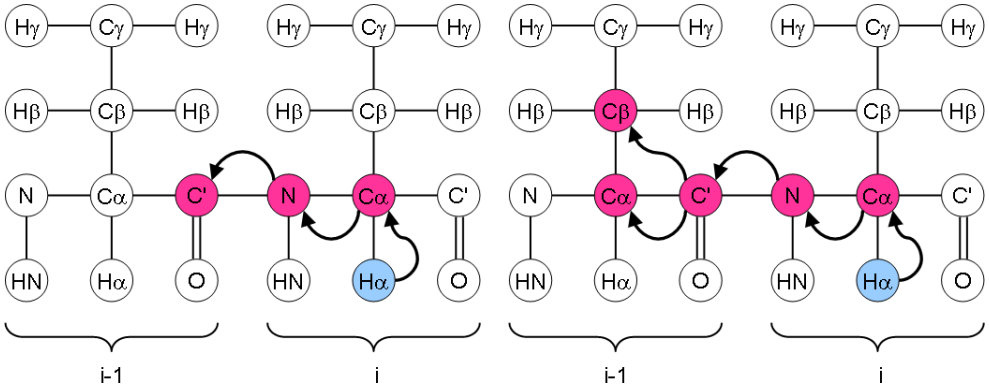


Fig 9. Observed magnetization transfers for a CANCO and CANCOCX protein detection scheme. Nuclei observed are in pink, nuclei through which magnetization flows are in light blue. Magnetization transfer at short mixing times in black arrows and at longer mixing times in grey arrows. Figure from www.protein-nmr.org.uk (10-2012).

3.6 CHHC and NHHC

The CHHC and NHHC[67] protein detection schemes are both “through-space” proton-proton protein detection schemes that provide structural restraints for protein structure calculations. The magnetization pathway starts from the hydrogen to the CA or N, then back to the hydrogen, through space to adjacent hydrogens and finally measured on the CA (fig 10).

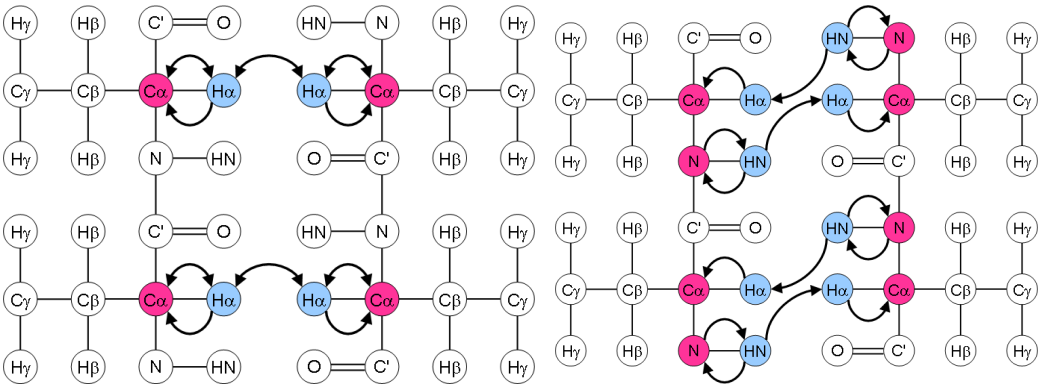


Fig 10. Observed magnetization transfers for a CHHC and NHHC protein detection scheme. Nuclei observed are in pink, nuclei through which magnetization flows are in light blue. Magnetization transfer is in black arrows. Figure from www.protein-nmr.org.uk (10-2012).

4. Discussion

There is no single answer to which labelling scheme and protein detection scheme should be used to unravel a protein structure. The strategy should be devised on the basis of the protein of interest and the available equipement. However some labelling schemes are good candidates to improve NMR spectra from certain protein detection schemes.

In general the TEASE (*t*en-*a*mino acid *s*elective and *e*xtensive labelling) labelling scheme, which specifically labels the amino acids produced from the glycolysis and PPP pathway, is a great labelling scheme for membrane proteins. The most hydrophobic amino acids are produced from the glycolysis and PPP pathway, thus labelling the membrane embedded segments predominantly. Furthermore the scrambling of labels due to the citric acid cycle can be neglected.

4.1 PDSD and DARR

The PDSD and DARR spectra are generally crowded for larger proteins; reduced labelling is potentially a great way to improve these spectra. This can range from reduced labelling by carbon sources or adding multiple labelled amino acids to the medium. The computational algorithm UPLABEL can be used[9]. Other combinational labelling (and reversed labelling) methods are also usable to optimize PDSD and DARR protein detection schemes[46].

4.2 NCA, NCO, NCACX, and NCOCX

By adding the additional PDSD or DARR step to the NCO or NCA scheme this the NCOCX and NCACX become three-dimensional. This greatly reduces overlap, however by reduced labelling this overlap problem can be further decreased.
Especially the reduced labelling by added carbon sources give specific labelling schemes on the CO and CA, which are essential for these protein detection schemes. By using 2,3-succinic acid, only the CO positions of Glu, Gln, and His are labelled; decreasing the spectral overlap immensely. Combining the assignments of 2,3-succinc acid with the assignments of 1-succinc acid should cover all the residues.
For glycerol or glucose the same strategy can be applied with slightly different CO positions labelled by 2-glycerol (Leu, half of the Asp, Asn, Met, and Thr, 66% of the Arg, Glu, Gln, and Pro).

4.3 CANCO with [1,3-^{13}C]-glycerol

The CANCO magnetization requires both a label on the CA and the CO positions. This greatly reduces the possibilities of reduced labelling schemes that can be used. Only 1,3-glycerol labels a few residues at both these positions, and only a part of these residues gets labelled at both positions.

20% Of the a-ketogluterates (Glu, Gln, Arg, and Pro) is expressed with a label on both the CA and CO, and 17% of the oxaloacetates (Asp, Asn, Met, Thr, and Ile) is expressed with a label on both the CA and CO (fig S1). Since the CANCO protein detection scheme has a low signal to noise ratio the residues labelled by these labelling schemes might give insufficient NMR signal.

4.4 CHHC and NHHC

Because these protein detection schemes involve “through-space” magnetization transfer, it is harder to predict which labelling schemes result in a reduction of spectral crowding. The CA is important in both these labelling schemes and reduced labelling schemes with specific CA labelling can simplify the assignment of the peaks on these NMR spectra.

5. List of references

1. Zheng, C., Han, L., Yap, C. W., Xie, B. & Chen, Y. Progress and problems in the exploration of therapeutic targets. *Drug discovery today* **11**, 412–20 (2006).

2. Ahram, M., Litou, Z. I., Fang, R. & Al-Tawallbeh, G. Estimation of membrane proteins in the human proteome. *In silico biology* **6**, 379–86 (2006).

3. Almén, M. S., Nordström, K. J. V, Fredriksson, R. & Schiöth, H. B. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC biology* **7**, 50 (2009).

4. Melén, K., Krogh, A. & Von Heijne, G. Reliability Measures for Membrane Protein Topology Prediction Algorithms. *Journal of Molecular Biology* **327**, 735–744 (2003).

5. Lee, A. *Lipid–protein interactions in biological membranes: a structural perspective*. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1612**, 1–40 (2003).

6. Huster, D., Yao, X., Jakes, K. & Hong, M. Conformational changes of colicin Ia channel-forming domain upon membrane binding: a solid-state NMR study. *Biochimica et biophysica acta* **1561**, 159–70 (2002).

7. Gonen, T., Sliz, P., Kistler, J., Cheng, Y. & Walz, T. Aquaporin-0 membrane junctions reveal the structure of a closed water pore. *Nature* **429**, 193–7 (2004).

8. Marassi, F. M. & Opella, S. J. NMR structural studies of membrane proteins. *Current opinion in structural biology* **8**, 640–8 (1998).

9. Hefke, F. *et al.* Optimization of amino acid type-specific 13C and 15N labeling for the backbone assignment of membrane proteins by solution- and solid-state NMR with the UPLABEL algorithm. *Journal of biomolecular NMR* **49**, 75–84 (2011).

10. Robinson, K. E., Reardon, P. N. & Spicer, L. D. Protein NMR Techniques. **831**, 261–277 (2012).

11. Serber, Z. & Dötsch, V. In-Cell NMR Spectroscopy. *Biochemistry* **40**, 14317–14323 (2001).

12. Reckel, S., Lopez, J. J., Löhr, F., Glaubitz, C. & Dötsch, V. In-cell solid-state NMR as a tool to study proteins in large complexes. *Chembiochem : a European journal of chemical biology* **13**, 534–7 (2012).

13. Fu, R. *et al.* In situ structural characterization of a recombinant protein in native Escherichia coli membranes with solid-state magic-angle-spinning NMR. *Journal of the American Chemical Society* **133**, 12370–3 (2011).

14. Higman, V. a *et al.* Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively 13C-labelled proteins. *Journal of biomolecular NMR* **44**, 245–60 (2009).

15. Goto, N., Gardner, K. & Mueller, G. A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *Journal of biomolecular …* **13**, 369–374 (1999).

16. Heise, H. *et al.* Molecular-level secondary structure, polymorphism, and dynamics of full-length α-synuclein fibrils studied by solid-state NMR. *Proceedings of the …* **102**, (2005).

17. Busche, A. E. L. *et al.* Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein. *Angewandte Chemie (International ed. in English)* **48**, 6128–31 (2009).

18. Campbell, E. a *et al.* Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* **104**, 901–12 (2001).

19. Serber, Z. *et al.* High-resolution macromolecular NMR spectroscopy inside living cells. *Journal of the American Chemical Society* **123**, 2446–7 (2001).

20. Serber, Z., Ledwidge, R., Miller, S. M. & Dötsch, V. Evaluation of parameters critical to observing proteins inside living Escherichia coli by in-cell NMR spectroscopy. *Journal of the American Chemical Society* **123**, 8895–901 (2001).

21. Almeida, F. C. *et al.* Selectively labeling the heterologous protein in Escherichia coli for NMR studies: a strategy to speed up NMR spectroscopy. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **148**, 142–6 (2001).

22. Cruzeiro-Silva, C., Albernaz, F. P., Valente, A. P. & Almeida, F. C. L. In-cell NMR spectroscopy. *Cell Biochemistry and Biophysics* **44**, 497–502 (2006).

23. Hong, M. Determination of Multiple [phi]-Torsion Angles in Proteins by Selective and Extensive 13C Labeling and Two-Dimensional Solid-State NMR. *Journal of Magnetic Resonance* **401**, 389–401 (1999).

24. Lundström, P. *et al.* Fractional 13C enrichment of isolated carbons using [1-13C]- or [2- 13C]-glucose facilitates the accurate measurement of dynamics at backbone Calpha and side-chain methyl positions in proteins. *Journal of biomolecular NMR* **38**, 199–212 (2007).

25. Gammeren, A. van Biosynthetic site-specific 13C labeling of the light-harvesting 2 protein complex: A model for solid state NMR structure determination of transmembrane proteins. *… of biomolecular NMR* 267–274 (2004).

26. Gammeren, a J. Van, Hulsbergen, F. B., Hollander, J. G. & Groot, H. J. M. De Residual backbone and side-chain 13C and 15N resonance assignments of the intrinsic transmembrane light-harvesting 2 protein complex by solid-state Magic Angle Spinning NMR spectroscopy. *Journal of biomolecular NMR* **31**, 279–93 (2005).

27. Glasky, a J. & Rafelson, M. E. The utilization of acetate-C14 by Escherichia coli grown on acetate as the sole carbon source. *The Journal of biological chemistry* **234**, 2118–22 (1959).

28. Hong, M. & Jakes, K. Selective and extensive 13C labeling of a membrane protein for solid-state NMR investigations. *Journal of biomolecular NMR* **14**, 71–4 (1999).

29. Filipp, F. V, Sinha, N., Jairam, L., Bradley, J. & Opella, S. J. Labeling strategies for 13C-detected aligned-sample solid-state NMR of proteins. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **201**, 121–30 (2009).

30. Senn, H. *et al.* Stereospecific assignment of the methyl lH NMR lines of valine and leucine in polypeptides by nonrandom 13C labelling. *FEBS letters* **249**, 113–118 (1989).

31. Nand, D., Cukkemane, A., Becker, S. & Baldus, M. Fractional deuteration applied to biomolecular solid-state NMR spectroscopy. *Journal of biomolecular NMR* **52**, 91–101 (2012).

32. Sibille, N. *et al.* Selective backbone labelling of ILV methyl labelled proteins. *Journal of biomolecular NMR* **43**, 219–27 (2009).

33. Mulder, F. a a, Hon, B., Mittermaier, A., Dahlquist, F. W. & Kay, L. E. Slow internal dynamics in proteins: application of NMR relaxation dispersion spectroscopy to methyl groups in a cavity mutant of T4 lysozyme. *Journal of the American Chemical Society* **124**, 1443–51 (2002).

34. Agarwal, V., Diehl, A., Skrynnikov, N. & Reif, B. High resolution 1H detected 1H,13C correlation spectra in MAS solid-state NMR using deuterated proteins with selective 1H,2H isotopic labeling of methyl groups. *Journal of the American Chemical Society* **128**, 12620–1 (2006).

35. Isaacson, R. L. *et al.* A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *Journal of the American Chemical Society* **129**, 15428–9 (2007).

36. Ayala, I., Sounier, R., Usé, N., Gans, P. & Boisbouvier, J. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *Journal of biomolecular NMR* **43**, 111–9 (2009).

37. Gelis, I. *et al.* Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* **131**, 756–69 (2007).

38. Velyvis, A., Ruschak, A. M. & Kay, L. E. An Economical Method for Production of 2 H , 13 CH 3 - Threonine for Solution NMR Studies of Large Protein Complexes : Application to the 670 kDa Proteasome. **7**, 1–8 (2012).

39. Takeuchi, K., Ng, E., Malia, T. J. & Wagner, G. 1-13C amino acid selective labeling in a 2H15N background for NMR studies of large proteins. *Journal of biomolecular NMR* **38**, 89–98 (2007).

40. Lin, M. T. *et al.* A rapid and robust method for selective isotope labeling of proteins. *Methods (San Diego, Calif.)* **55**, 370–8 (2011).

41. Tong, K. I., Yamamoto, M. & Tanaka, T. A simple method for amino acid selective isotope labeling of recombinant proteins in E. coli. *Journal of biomolecular NMR* **42**, 59–67 (2008).

42. Tong, K. I., Yamamoto, M. & Tanaka, T. Intrinsically Disordered Protein Analysis. **896**, 439–448 (2012).

43. Krishnarjuna, B., Jaipuria, G., Thakur, A., D'Silva, P. & Atreya, H. S. Amino acid selective unlabeling for sequence specific resonance assignments in proteins. *Journal of biomolecular NMR* **49**, 39–51 (2011).

44. Etzkorn, M. *et al.* Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy. *Angewandte Chemie (International ed. in English)* **46**, 459–62 (2007).

45. Schneider, R. *et al.* Solid-state NMR spectroscopy applied to a chimeric potassium channel in lipid bilayers. *Journal of the American Chemical Society* **130**, 7427–35 (2008).

46. Hiroaki, H. & Umetsu, Y. A simplified recipe for assigning amide NMR signals using combinatorial 14 N amino acid inverse-labeling. 167–174 (2011).doi:10.1007/s10969-011-9116-0

47. Jaipuria, G., Krishnarjuna, B., Mondal, S., Dubey, A. & Atreya, H. S. Isotope labeling in Biomolecular NMR. **992**, (2012).

48. Young, T. S. & Schultz, P. G. Beyond the canonical 20 amino acids: expanding the genetic lexicon. *The Journal of biological chemistry* **285**, 11039–44 (2010).

49. Cellitti, S. E. *et al.* In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *Journal of the American Chemical Society* **130**, 9268–81 (2008).

50. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annual review of biochemistry* **79**, 413–44 (2010).

51. Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annual review of biophysics and biomolecular structure* **35**, 225–49 (2006).

52. Muralidharan, V. & Muir, T. W. Protein ligation : an enabling technology for the biophysical analysis of proteins. **3**, 429–438 (2006).

53. Giriat, I. & Muir, T. W. Protein semi-synthesis in living cells. *Journal of the American Chemical Society* **125**, 7180–1 (2003).

54. Ong, S.-E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature protocols* **1**, 2650–60 (2006).

55. Ong, S.-E. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics* **1**, 376–386 (2002).

56. Gu, S., Pan, S., Bradbury, E. M. & Chen, X. Precise Peptide Sequencing and Protein Quantification in the Human Proteome. **0305**, (2003).

57. Ong, S., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. **1**, 1–8 (2004).

58. Ong, S., Kratchmarova, I. & Mann, M. C-Substituted Arginine in Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) research articles. 173–181 (2003).

59. Ibarrola, N., Molina, H., Iwahori, A. & Pandey, A. A novel proteomic approach for specific identification of tyrosine kinase substrates using [13C]tyrosine. *The Journal of biological chemistry* **279**, 15805–13 (2004).

60. Bloembergen, N. On the interaction of nuclear spins in a crystalline lattice. *Physica* **15**, 386–426 (1949).

61. Takegoshi, K., Nakamura, S. & Terao, T. 13C-1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chemical Physics Letters* **344**, 631–637 (2001).

62. Baldus, M., Petkova, A., Herzfeld, J. & Griffin, R. Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. *Molecular Physics* **95**, 1197–1207 (1998).

63. Pauli, J., Baldus, M., Van Rossum, B., De Groot, H. & Oschkinat, H. Backbone and side-chain 13C and 15N signal assignments of the alpha-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 Tesla. *Chembiochem : a European journal of chemical biology* **2**, 272–81 (2001).

64. Verel, R., Ernst, M. & Meier, B. H. Adiabatic dipolar recoupling in solid-state NMR: the DREAM scheme. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **150**, 81–99 (2001).

65. Li, Y., Berthold, D. a, Frericks, H. L., Gennis, R. B. & Rienstra, C. M. Partial (13)C and (15)N chemical-shift assignments of the disulfide-bond-forming enzyme DsbB by 3D magic-angle spinning NMR spectroscopy. *Chembiochem : a European journal of chemical biology* **8**, 434–42 (2007).

66. Franks, W. T., Kloepper, K. D., Wylie, B. J. & Rienstra, C. M. Four-dimensional heteronuclear correlation experiments for chemical shift assignment of solid proteins. *Journal of biomolecular NMR* **39**, 107–31 (2007).

67. Lange, A., Luca, S. & Baldus, M. Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids. *Journal of the American Chemical Society* **124**, 9704–5 (2002).

6. Supplementary figures

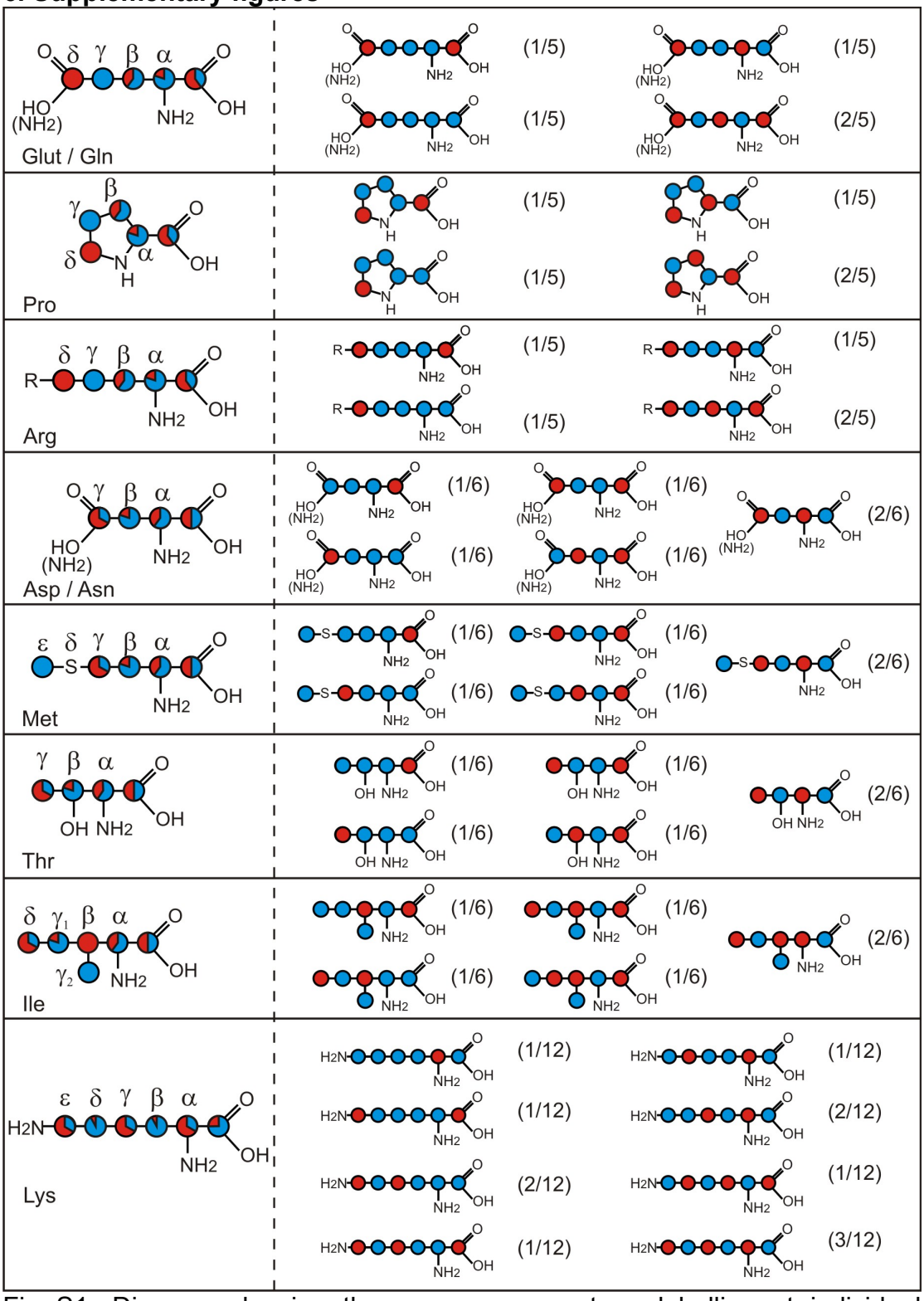


Fig S1. Diagram showing the average percentage labelling at individual atomic sites (left) and the isotopomer populations (right) for the amino acids synthesised via the citric acid pathway, as observed in solution NMR spectra of the chicken a-spectrin SH3 domain (Castellani et al. 2002). Sites which are ^{13}C labelled by growth on [1,3-^{13}C]-glycerol are shown in blue; sites labelled by growth on [2-^{13}C]-glycerol are shown in red. Figure from Higman et al.

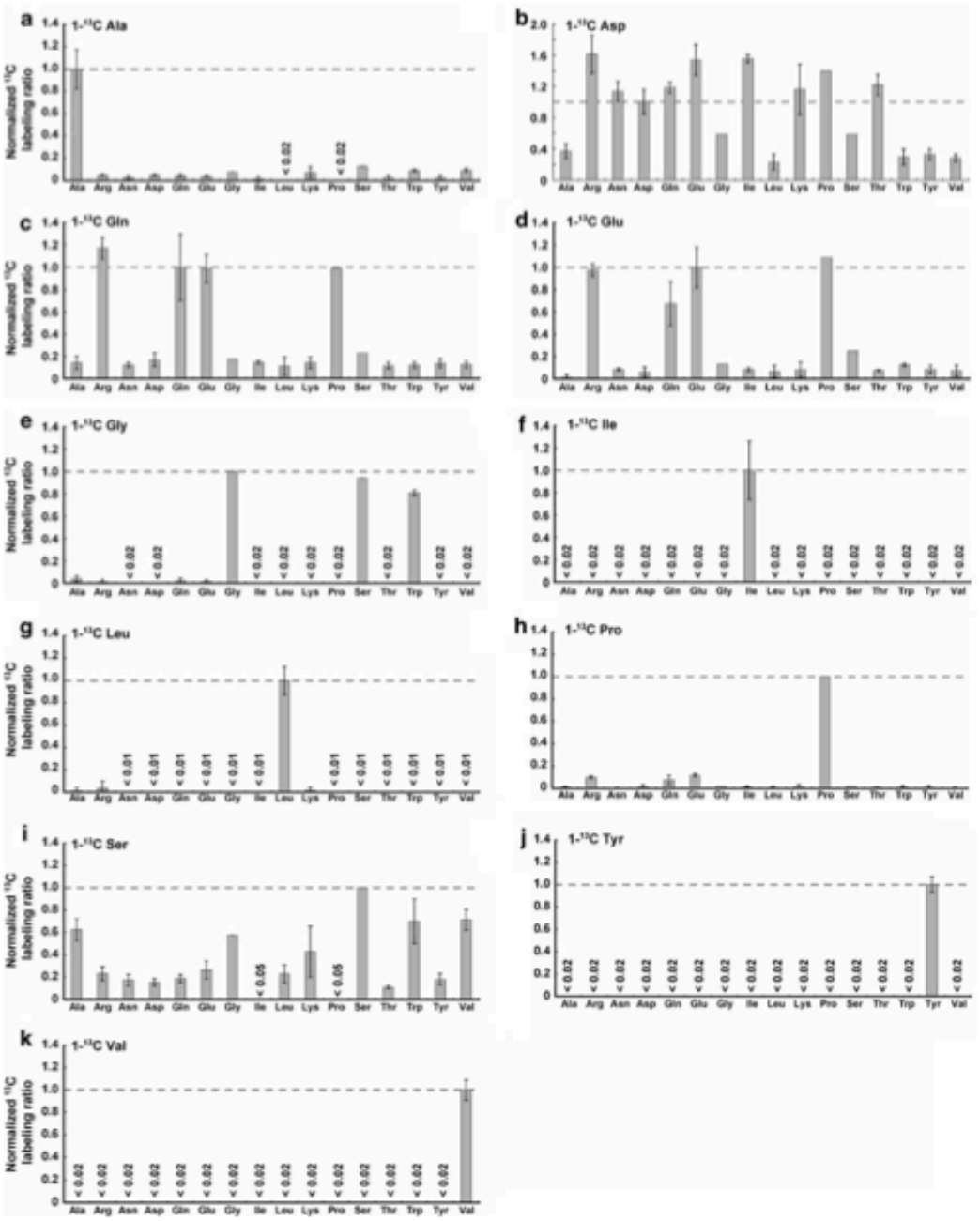


Fig S2. Amino acid specific ^{13}C CO scrambling. Figure from Takeuchi et al.

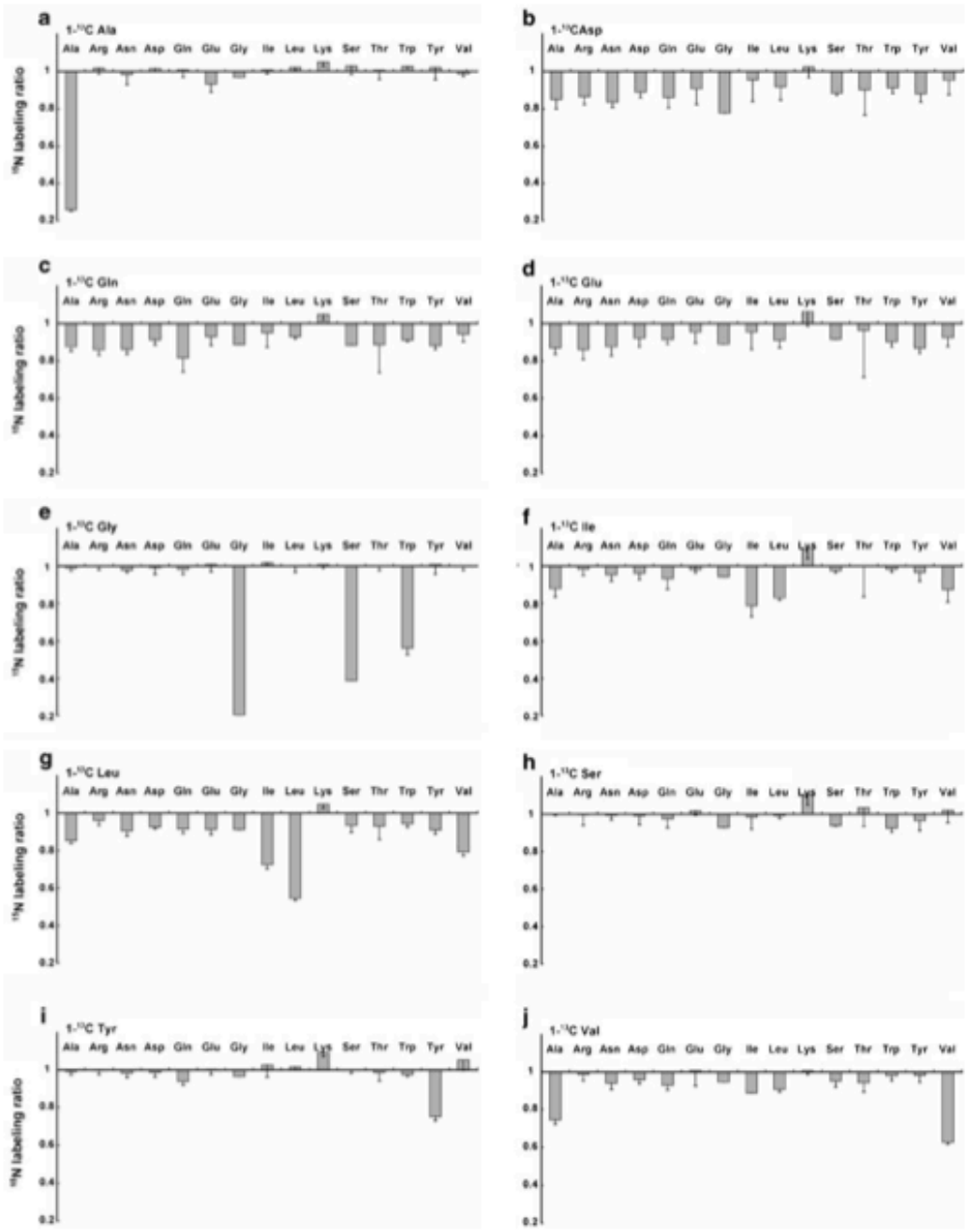


Fig S3. Amino acid specific ^{15}N scrambling. Fig from Takeuchi et al.